

Komparasi Algoritma *Support Vector Machine* dan *Naïve Bayes* pada Klasifikasi Jenis Buah Kurma berdasarkan Citra *Hue Saturation Value*

Comparison of Support Vector Machine and Naïve Bayes Algorithms on Date Fruit Type Classification based on Hue Saturation Value Image

¹Mia Risa Br. Keliat*, ²Muhammad Ikhsan

^{1,2}Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara

^{1,2}Jl. Lapangan Golf, Desa Durin Jangak, Kec. Pancur Batu, Kab. Deli Serdang, Sumatera Utara, Indonesia

*e-mail: mia0701202194@uinsu.ac.id

(received: 11 January 2025, revised: 14 January 2025, accepted: 15 January 2025)

Abstrak

Buah kurma merupakan salah satu buah yang populer di Indonesia, terutama selama bulan Ramadan. Dengan meningkatnya konsumsi, diperlukan sistem otomatis untuk mengklasifikasi jenis kurma guna menghindari kesalahan dalam identifikasi. Metode yang digunakan adalah dua algoritma klasifikasi yang berbeda yaitu SVM dan Naïve Bayes. Penelitian ini membandingkan algoritma *Support Vector Machine* (SVM) dan Gaussian Naïve Bayes dalam klasifikasi jenis buah kurma berdasarkan citra Hue, Saturation, Value (HSV). Dataset terdiri dari 200 citra kurma dari empat jenis, yaitu Kurma Ajwa, Kurma Sukari, *Kurma Golden Valley*, dan Kurma *Deglet Nour*, dengan pembagian 70% untuk data latih dan 30% untuk data uji dan data diambil menggunakan kamera *smartphone* beresolusi tinggi dalam kondisi pencahayaan terkendali untuk menjaga konsistensi kualitas gambar. Proses segmentasi mencakup dari RGB (asli), greyscale, biner, Operasi Komplemen, *Filling Holes* dan konversi ke ruang warna HSV. *Tools* matlab digunakan untuk implementasi algoritma dan evaluasi performa model. Hasil penelitian menunjukkan bahwa Gaussian Naïve Bayes memiliki kinerja yang lebih baik dengan akurasi 80,00%, *precision* 79,74%, *recall* 78,46%, dan *F1-score* 79,09%. Sebaliknya, SVM dengan kernel linier hanya mencapai akurasi 66,67%, *precision* 52,49%, *recall* 65,00%, dan *F1-score* 58,08%. Evaluasi menunjukkan bahwa kedua model tidak mengalami *overfitting*. Berdasarkan analisis GUI, Naïve Bayes terbukti lebih unggul dalam klasifikasi jenis kurma. Penelitian ini memberikan kontribusi signifikan pada pengembangan sistem otomatis berbasis citra untuk klasifikasi produk agrikultur.

Kata kunci: buah kurma, citra HSV, matlab, naïve bayes, support vector machine (SVM)

Abstract

Dates are a popular fruit in Indonesia, especially during the month of Ramadan. With the increase in consumption, an automated system is needed to classify different types of dates to avoid misidentification. This study compares two classification algorithms: Support Vector Machine (SVM) and Naïve Bayes. The research compares the performance of the Support Vector Machine (SVM) and Gaussian Naïve Bayes algorithms in classifying date fruits based on Hue, Saturation, Value (HSV) images. The dataset consists of 200 images of dates from four types: Ajwa Dates, Sukari Dates, Golden Valley Dates, and Deglet Nour Dates, with a 70% training data and 30% testing data split. The images were captured using a high-resolution smartphone camera under controlled lighting conditions to ensure consistent image quality. The segmentation process includes converting from RGB (original), grayscale, binary, complement operation, filling holes, and conversion to the HSV color space. MATLAB tools were used to implement the algorithms and evaluate model performance. The results show that Gaussian Naïve Bayes outperforms the SVM with a higher accuracy of 80.00%, precision of 79.74%, recall of 78.46%, and F1-score of 79.09%. In contrast, the SVM with a linear kernel only achieved an accuracy of 66.67%, precision of 52.49%, recall of 65.00%, and F1-score of

<http://sistemasi.ftik.unisi.ac.id>

58.08%. Evaluation showed that neither model suffered from overfitting. Based on the GUI analysis, Naïve Bayes proved superior in classifying the types of dates. This study makes a significant contribution to the development of an automated image-based system for classifying agricultural products.

Keywords: : dates, HSV image, matlab, naïve bayes, support vector machine (SVM)

1 Pendahuluan

Buah kurma merupakan salah satu buah yang sangat populer di Indonesia terlebih lagi pada saat bulan Ramadhan. Ada banyak sekali jenis buah kurma yang ada dipasaran. Menilik data Badan Pusat Statistik (BPS), pada tahun 2022 impor kurma Indonesia mencapai angka 61,35 ribu ton dengan nilai sebesar US\$ 86,26 juta atau sekitar Rp 1,3 triliun. Angka ini naik 22,38% dibandingkan tahun sebelumnya. Oleh karena itu, klasifikasi jenis buah kurma harus dilakukan agar masing-masing jenis buah kurma tersebut tidak saling tertukar satu sama lain [1]. Di Indonesia, banyak jenis kurma yang dijual, namun jenis kurma yang paling diminati termasuk kurma Ajwa, kurma Sukari, kurma *Deglet Nour* dan kurma *Golden Valley* [2]. Kurma juga mengandung riboflavin, niasin, piridoksal, dan folat dimana dalam 100 gram kurma memenuhi lebih dari 9% kebutuhan vitamin dalam sehari. Buah kurma yang telah matang juga kaya akan kandungan kalsium dan zat besi [3].

SVM merupakan salah satu algoritma klasifikasi yang kuat dan sering digunakan dalam berbagai aplikasi pemrosesan citra. Algoritma SVM bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan kelas-kelas dalam ruang fitur [4]. SVM cenderung memiliki kinerja yang baik dalam kasus di mana terdapat margin kejelasan yang tinggi antara kelas-kelas yang akan dipisahkan [5].

Menurut Ellif [6] salah satu algoritma dalam teknik klasifikasi adalah algoritma Naive Bayes. Naive Bayes adalah klasifikasi ilmuwan Inggris yang menggunakan metode probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman masa lalu. Ini dikenal sebagai Teorema Bayes dan dikombinasikan dengan Naive, yang menganggap kondisi antar atribut saling bebas. Menurut klasifikasi Naive Bayes, apakah ada atau tidak karakteristik tertentu yang membentuk kelas tertentu tidak terkait dengan karakteristik kelas lainnya. Naive Bayes adalah algoritma klasifikasi yang berbasis pada teorema Bayes dengan asumsi naif, yaitu bahwa semua fitur dalam data adalah independen satu sama lain ketika diberikan kelas target. Meskipun sederhana, Naive Bayes sering kali efektif dalam klasifikasi teks dan data dengan fitur-fitur yang bersifat kategorikal.

Dalam konteks klasifikasi buah kurma berdasarkan citra HSV, kedua algoritma ini dapat diterapkan dengan beberapa pertimbangan dimana SVM dapat bekerja dengan baik jika terdapat pola-pola kompleks dalam data yang memerlukan pemisahan yang rumit. SVM mampu menangani ruang fitur yang besar dan dapat menangani data yang tidak terstruktur dengan baik [4]. Naive Bayes mungkin lebih cocok jika fitur-fitur dalam citra sudah direduksi menjadi fitur-fitur yang bersifat independen, misalnya dengan menggunakan metode ekstraksi fitur yang sesuai. Jika asumsi independensi antar-fitur cukup berlaku, Naive Bayes dapat memberikan hasil yang baik dengan waktu komputasi yang relatif rendah. Namun, dalam prakteknya, pemilihan antara SVM dan Naive Bayes tergantung pada karakteristik data, kompleksitas masalah, serta tujuan dari sistem klasifikasi buah kurma berdasarkan citra HSV tersebut.

Tujuan dari komparasi algoritma *Support Vector Machine* (SVM) dan Naive Bayes dalam klasifikasi jenis buah kurma berdasarkan citra Hue, Saturation, dan Value (HSV) adalah untuk memberikan pemahaman yang lebih baik tentang aspek kinerja, akurasi dari masing-masing algoritma dalam konteks yang spesifik dan menganalisis hasil akurasi, *precision*, dan *recall* dari setiap algoritma untuk menentukan algoritma yang paling optimal. Penelitian ini memberikan manfaat dalam pengembangan metodologi klasifikasi buah kurma dengan menggunakan fitur citra HSV. Metodologi ini dapat digunakan sebagai referensi dalam pengembangan sistem klasifikasi berbasis citra untuk jenis produk atau objek lainnya.

2 Tinjauan Literatur

Terdapat beberapa penelitian yang telah dilakukan dan berkaitan dengan penelitian ini. Adapun pada penelitian Ellif [6] tentang Klasifikasi Kematangan Pepaya Menggunakan Ruang Warna HSV dan Metode Naïve Bayes Classifier, peneliti menggunakan sampel kematangan buah pepaya dengan

50 data buah pepaya dibagi menjadi 30 data latih dan 20 data uji menggunakan ekstraksi warna HSV dengan metode naïve bayes dapat diimplementasikan dengan baik. Dan aplikasi yang dibangun menggunakan model warna HSV dengan metode naïve bayes didapatkan akurasi 100%.

Penelitian yang dilakukan Muhammad [7] dengan judul Klasifikasi Tingkat Kematangan Buah Pisang Berdasarkan Fitur Warna dengan Metode SVM. Penelitian ini menggunakan dataset berupa 80 buah citra pisang diambil dari situs Kaggle. Data penelitian menggunakan fitur ekstraksi untuk menentukan tingkat kematangan melalui fitur warna yaitu RGA. Dari fitur tersebut dilakukan pengujian klasifikasi menggunakan metode SVM, setelah dilakukan pengujian didapat akurasi kematangan sebesar 75% dari 80 data citra pisang yang diuji.

Pada penelitian Ayuningsih [8] dengan judul Klasifikasi Citra Makanan Menggunakan HSV *Color Moment* dan *Local Binary Pattern* dengan Naïve Bayes *Classifier*. Nilai akurasi menggunakan metode LBP dan HSV menghasilkan nilai sebesar 65%. Selain itu, hasil dari pengujian yang telah dilakukan menggunakan metode HSV menghasilkan akurasi sebesar 65% dan metode LBP menghasilkan akurasi sebesar 60%.

Menurut penelitian Milda [9] dengan judul Klasifikasi Kematangan Buah Pisang Kepok Berdasarkan Citra HSV dengan K-Nearest Neighbors. Pada sistem klasifikasi tingkat kematangan buah pisang kepok menggunakan algoritma KNN memiliki akurasi tertinggi pada nilai $K=8$. Penggunaan ekstraksi fitur HSV dan algoritma KNN menghasilkan akurasi 98% dalam mengklasifikasikan tingkat kematangan buah pisang kepok.

Pada penelitian Afriansyah [10] dengan judul Optimasi Algoritma Naïve Bayes pada Klasifikasi Buah Apel Berdasarkan Fitur Warna RGB mendapatkan nilai rata-rata akurasi sebesar 88.88% dengan menggunakan 150 citra buah apel berdasarkan tingkat kematangan apel. Pada penelitian ini digunakan *K-fold cross validation* sebagai metode yang digunakan untuk membagi dataset menjadi sejumlah k buah partisipasi secara acak dengan jumlah $k=3$.

Pada penelitian Saputra [11] dengan judul Klasifikasi Tingkat Kemanisan Alpukat Berdasarkan Fitur *Hue Saturation Value* (HSV) dengan Menggunakan *Support Vector Machine* (SVM) mendapatkan hasil akurasi sebesar 100% menggunakan kernel *polynomial* dan 98.91% menggunakan kernel *Gaussian*. Pembagian dataset citra untuk data uji dan data latih digunakan *Fold Cross Validation* jenis *Leave One Out, fold* yang digunakan yaitu *4-fold*, *5-fold*, serta *10-fold*.

Pada penelitian Setyawan [12] dengan judul Impelentasi *Support Vector Machine* (SVM) pada Klasifikasi Jenis Tanah Memanfaatkan Fitur RGB mendapatkan akurasi sebesar 88% menggunakan 400 citra dari 4 jenis tanah yaitu *clay soil*, *black soil*, *red soil*, dan *alluvial soil*. Dataset pada penelitian ini diubah ukuran dan formatnya menjadi 512x512 piksel untuk mengurangi beban komputasi, serta hasil klasifikasi yang akurat.

Pada penelitian Afriansyah [13] dengan judul Algoritma Naïve Bayes yang Efisien untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna mendapatkan rata-rata akurasi sebesar 86.66% dari 3 kategori buah pisang raja dan menggunakan 180 citra buah pisang raja. *K-fold cross validation* sebagai metode yang digunakan untuk membagi dataset. Jumlah k yang digunakan adalah 3 yaitu manis, sepat, dan hambar.

Adapun perbedaan antara penelitian ini dengan penelitian sebelumnya adalah objek penelitian, pendekatan penelitian, serta menggunakan dua algoritma berbeda yang akan dibandingkan hasil akurasinya. Penelitian ini membahas tentang algoritma SVM linear dan Naïve Bayes *gaussian* pada citra HSV yang sebelumnya sudah dilakukan proses segmentasi, dari citra RGB diubah ke *greyscale*, lalu diubah ke citra biner, setelah itu dilakukan operasi komplemen, dilanjutkan dengan *filling holes* dan terakhir citra HSV. Sedangkan penelitian sebelumnya membahas berbagai objek, jenis kernel yang digunakan, serta *pre-processing* citra yang berbeda dengan penelitian ini.

Salah satu algoritma pembelajaran mesin yang sangat baik untuk mengklasifikasikan data adalah *Support Vector Machine* (SVM). SVM juga dikenal sebagai *classifier linier* yang didasarkan pada prinsip memaksimalkan margin. Secara optimal, SVM mengklasifikasikan data menjadi dua kelompok dalam ruang dimensi yang lebih besar dengan menggunakan *hyperplane*. Margin adalah jarak antara data terdekat dari setiap kelas dan *hyperplane*. *Hyperplane* adalah pemisah terbaik antara dua kelas yang telah ditentukan, dan data terdekat ini disebut support vektor. Prinsip dasar SVM adalah pengklasifikasi linier; kemudian dikembangkan untuk menangani masalah non-linier, seperti konsep trik kernel di ruang kerja besar [5].

Menurut Saleh [14] Teorema Bayes digunakan dalam algoritma untuk memperkirakan semua atribut yang bebas dan saling lepas yang dapat diberikan oleh suatu nilai pada kelas variabel. Teorema Bayes adalah klasifikasi kemungkinan sederhana yang dapat menghitung seluruh kemungkinan dengan menggabungkan sejumlah kombinasi dan frekuensi dari nilai yang diperoleh dari basis data. Naive Bayes adalah klasifikasi yang menggunakan metode kemungkinan dan perhitungan yang ditemukan oleh Thomas Bayes, seorang ilmuwan Inggris, untuk memprediksi peluang yang akan datang berdasarkan pengalaman sebelumnya.

Penelitian ini diharapkan dapat memberikan kontribusi kepada peneliti berikutnya yang mungkin menggunakan Algoritma *Support Vector Machine* dan *Naïve Bayes* pada penelitiannya. Penelitian ini memberikan kontribusi dalam pengembangan metodologi klasifikasi buah kurma dengan menggunakan fitur citra HSV. Metodologi ini dapat digunakan sebagai referensi dalam pengembangan sistem klasifikasi berbasis citra untuk jenis produk atau objek lainnya. Bagi pembaca penelitian ini dapat memberikan kontribusi dalam pemilihan jenis buah kurma yang sesuai.

3 Metode Penelitian

Metodologi penelitian merupakan panduan untuk mendapatkan data dan informasi dalam menyelesaikan masalah yang akan dibahas juga untuk membantu peneliti dalam menyusun kerangka beserta tahapan penelitian. Adapun tahapan yang harus dilalui untuk melakukan penelitian ini yaitu sebagai berikut.

3.1 Pendekatan penelitian

Pendekatan penelitian yang digunakan dalam penelitian ini adalah pendekatan eksperimental, yang menguji hipotesis melalui analisis statistik pada beberapa variabel yang terlibat. Serangkaian uji khusus diperlukan untuk menjelaskan dan mengukur variabel tersebut dengan masalah penelitian [15]. Pendekatan ini dipilih karena tujuan penelitian ini untuk membandingkan atau mengkomparasi dua algoritma yaitu SVM dan NB pada klasifikasi jenis buah kurma berdasarkan HSV. Dengan menggunakan pendekatan eksperimental, peneliti dapat mengontrol variabel-variabel lain yang dapat mempengaruhi hasil penelitian, seperti kualitas citra, ukuran dataset, dan parameter algoritma. Pendekatan eksperimental memungkinkan peneliti untuk mengumpulkan data citra HSV dari berbagai jenis buah kurma dan kemudian menggunakan algoritma SVM dan NB untuk mengklasifikasi jenis buah kurma yang terkait dengan setiap citra. Dengan demikian, peneliti dapat melakukan analisis yang lebih detail dan akurat dari kinerja algoritma SVM dan NB dalam klasifikasi jenis buah kurma berdasarkan citra HSV. Metodologi penelitian yang akan digunakan meliputi tahapan pengumpulan data, *pre-processing* data, ekstraksi fitur HSV, implementasi algoritma SVM, implementasi NB, evaluasi dan analisis hasil.

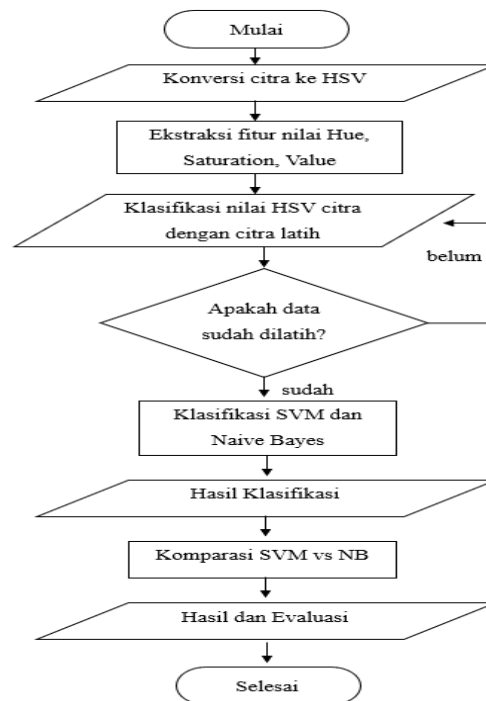
3.2 Pengumpulan data

Pada tahap pengumpulan data, tahap ini dilakukan dengan mengumpulkan data berdasarkan 4 jenis buah kurma yang mana masing-masing jenisnya menggunakan 200 citra buah kurma. Sebelum itu, dilakukan juga wawancara dan observasi guna mengetahui jenis kurma yang baik dan benar. Setelah melakukan observasi dan wawancara data citra kurma akan diambil secara langsung menggunakan kamera *smartphone* yang nantinya akan digunakan sebagai data latih dan data uji pada penelitian ini. Adapun data citra untuk data latih pada penelitian ini menggunakan sebanyak 200 citra gambar yang terdiri dari 4 jenis buah kurma. Sedangkan untuk data uji akan digunakan sebanyak 160 citra gambar, terdiri dari 40 kurma sukari, 40 kurma ajwa, 40 kurma *golden valley* dan 40 kurma *deglet nour*. Pada data uji ini objek penelitian akan langsung diambil gambarnya secara *realtime* ketika sistem dijalankan dengan jumlah data *testing*.

Citra kurma yang diperoleh memiliki resolusi tinggi, sehingga dilakukan proses penyesuaian ukuran untuk menyamakan dimensi citra, yaitu sebesar 128x128 piksel. Langkah ini dilakukan untuk memastikan konsistensi data selama proses pengolahan citra dan ekstraksi fitur. Setelah itu, dataset dikelompokkan ke dalam empat folder berdasarkan jenis kurma, yaitu Kurma_Ajwa, Kurma_Deglet, Kurma_Golden, dan Kurma_Sukari. Pengelompokan ini bertujuan untuk mempermudah proses klasifikasi pada tahap selanjutnya.

3.3 Rencana pembahasan

Perancangan yaitu tahap dari perencanaan, penggambaran dan pembuatan sketsa. Penelitian ini akan diproses menggunakan Matlab versi 2016. Serta penelitian ini akan dilakukan dengan melalui beberapa tahapan dimulai dengan tahapan pengumpulan data yang dibutuhkan hingga melalui proses rancangan model dari algoritma SVM dan NB berdasarkan Hue Saturatin Value. Berdasarkan pada Gambar 1 dibawah alur penelitian ini dimulai dengan menyiapkan citra asli (RGB) yang nantinya akan dikonversi ke citra HSV dan dihitung setiap nilai Hue, Saturation, dan Value. Selanjutnya citra akan diklasifikasi apakah sudah dilatih atau belum, jika sudah citra akan lanjut melakukan proses klasifikasi untuk algoritma SVM dan Naïve Bayes. Hasil klasifikasi nantinya akan dikomparasi dan dieva



Gambar 1. Flowchart penelitian

4 Hasil dan Pembahasan

Pada bagian ini dijelaskan mengenai dataset dan pembagian data uji dan data latih yang digunakan pada penelitian ini. Pada bagian ini juga dijelaskan bagaimana proses hitung manual penerapan algoritma Support Vector Machine dan Naive Bayes.

4.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini terdiri dari 200 citra buah kurma yang telah dikelompokkan berdasarkan empat jenis, yaitu Kurma Ajwa sebanyak 55 data, Kurma Deglet sebanyak 50 data, Kurma Golden sebanyak 50, dan Kurma Sukari sebanyak 45. Setiap jenis kurma memiliki citra dengan variasi sudut pengambilan gambar dan kondisi permukaan buah. Dataset ini disimpan dalam folder terpisah untuk mempermudah proses segmentasi dan klasifikasi.

Setiap citra pada dataset direpresentasikan dalam format warna RGB (*Red, Green, Blue*). Namun, untuk kebutuhan penelitian, citra ini dikonversi ke ruang warna HSV (*Hue, Saturation, Value*). Ekstraksi fitur dilakukan dengan menghitung rata-rata nilai Hue, Saturation, dan Value pada setiap citra, sehingga menghasilkan vektor fitur berukuran tiga dimensi untuk masing-masing gambar. Fitur-fitur ini digunakan sebagai masukan pada algoritma klasifikasi.

Berikut contoh perhitungan citra ke HSV dimana data RGB yang didapat adalah sebagai berikut:

Piksel 1 : [145,100,75]
Piksel 2 : [200,180,150]
Piksel 3 : [255,200,100]
Piksel 4 : [100,50,200]

Piksel 5 : [180,180,180]

Piksel 6 : [50,100,75]

Piksel 7 : [200,100,150]

Piksel 8 : [0,255,255]

Piksel 9 : [255,0,0]

Piksel 10 : [0,0,255]

Langkah awal adalah normalisasi misalnya untuk piksel 1 (R = 145, G = 100, B = 75):

$$R' = \frac{145}{255} \approx 0.569, \quad G' = \frac{100}{255} \approx 0.392, \quad B' = \frac{75}{255} \approx 0.294$$

Kemudian menghitung C_{max} , C_{min} , dan Δ :

$$C_{max} = \max(0.569, 0.392, 0.294) = 0.569$$

$$C_{min} = \min(0.569, 0.392, 0.294) = 0.294$$

$$\Delta = C_{max} - C_{min} = 0.569 - 0.294 = 0.275$$

Menghitung nilai H (Hue):

$$H = 60^\circ * \left(\frac{G' - B'}{\Delta} \text{ mod } 6 \right) = 60^\circ * \left(\frac{0.392 - 0.294}{0.275} \text{ mod } 6 \right)$$

$$H = 60^\circ * 0.356 = 21.4^\circ$$

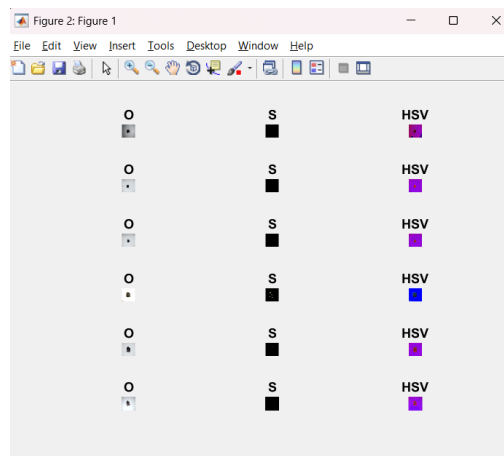
Menghitung S (Saturation):

$$S = \frac{\Delta}{C_{max}} = \frac{0.275}{0.569} = 0.483$$

Menghitung V (Value):

$$V = C_{max} = 0.569$$

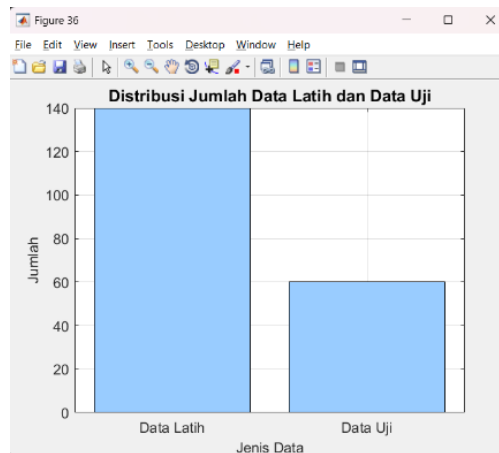
Sehingga dapat disimpulkan nilai HSV untuk piksel 1 yaitu $H = 21.4^\circ$, $S=0.483$ dan $V=0.569$. Seperti pada Gambar 2 dibawah dimana O merupakan citra Original (Asli), S merupakan citra yang telah dilakukan segmentasi, dan HSV merupakan citra yang telah dilakukan ekstrasi ciri menjadi citra HSV.



Gambar 2. Output citra asli > segmentasi > HSV

4.2 Pembagian Data Uji dan Data Latih

Pada Gambar 3 dapat dilihat pembagian jumlah data latih dan data uji, dimana data latih terdiri dari 140 citra dan data uji 60 citra.



Gambar 3. Distribusi jumlah data uji dan data latih

Dataset terdiri dari empat kategori, yaitu Kurma Ajwa, Kurma Deglet, Kurma Golden, dan Kurma Sukari. Proses Label Encoding memberikan representasi numerik sebagai berikut:

1. Kurma Ajwa : 1
2. Kurma Deglet : 2
3. Kurma Golden : 3
4. Kurma Sukari : 4

Model dilatih dengan metode cross-validation lima kali lipat (5-fold cross-validation) untuk menghindari overfitting dan memastikan performa yang konsisten. Seperti yang terlihat pada Tabel 1 dan Tabel 2 dimana digunakan 8 data latih dan 4 data uji sebagai contoh perhitungan Algoritma SVM dan Naive Bayes.

Tabel 1. Data latih

Mean_H	Mean_S	Mean_V	Std_H	Std_S	Std_V	Label
0.551266	0.037152	0.559529	0.141054	0.03064	0.161347	1
0.116428	0.158872	0.600096	0.027029	0.191659	0.140134	2
0.106275	0.179755	0.659874	0.016188	0.162941	0.164733	3
0.530638	0.057437	0.802345	0.136108	0.11606	0.15211	4
0.58494	0.038561	0.591155	0.100915	0.032327	0.166843	1
0.55642	0.115003	0.826382	0.170252	0.160718	0.171746	2
0.106275	0.179755	0.659874	0.016188	0.162941	0.164733	3
0.508713	0.098443	0.758231	0.180642	0.186961	0.188316	4

Tabel 2. Data uji

Mean_H	Mean_S	Mean_V	Std_H	Std_S	Std_V	Label
0.578192	0.03728	0.816108	0.043001	0.024703	0.135981	1
0.13406	0.095925	0.921496	0.066812	0.204569	0.200082	2
0.113857	0.165211	0.591572	0.028634	0.163849	0.158248	3
0.49869	0.078108	0.785001	0.186981	0.152674	0.184682	4

4.3 Penerapan Algoritma SVM

Sebelum melakukan prediksi terhadap data uji menggunakan model SVM linier, penting untuk memahami bahwa fungsi keputusan seperti pada Persamaan (1) berikut:

$$f(x) = w \cdot x + b \quad (1)$$

Dimana:

w : vektor yang tegak lurus dengan *hyperplane*

x : data

b : nilai bias

$f(x)$: fungsi *hyperplane*

Berdasarkan data latih, kita telah menghitung nilai $w = [0.2175, -0.0609]$ dan $b = 0.8824$. Selanjutnya, kita akan menggunakan fungsi ini untuk menghitung nilai $f(x)$ untuk setiap sampel data uji.

Kelas 1 (Label 1) :

$$X_1 = \begin{bmatrix} 0.5513 & 0.0372 \\ 0.5849 & 0.0345 \end{bmatrix}$$

Kelas 2 (Label -1)

$$X_2 = \begin{bmatrix} 0.1164 & 0.1589 \\ 0.5564 & 0.1105 \end{bmatrix}$$

$$\text{Label : } Y = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

$$\text{Kernel Linier : } K(x_i, x_j) = x_i \cdot x_j \quad (2)$$

Hitung elemen matriks kernel untuk pasangan data berdasarkan Persamaan 2 :

$$K = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & x_1 \cdot x_3 & x_1 \cdot x_4 \\ x_2 \cdot x_1 & x_2 \cdot x_2 & x_2 \cdot x_3 & x_2 \cdot x_4 \\ x_3 \cdot x_1 & x_3 \cdot x_2 & x_3 \cdot x_3 & x_3 \cdot x_4 \\ x_4 \cdot x_1 & x_4 \cdot x_2 & x_4 \cdot x_3 & x_4 \cdot x_4 \end{bmatrix}$$

Perhitungan Elemen :

$$\begin{aligned} x_1 \cdot x_1 &= (0.5513 * 0.5513) + (0.0372 * 0.0372) = 0.3048 \\ x_1 \cdot x_2 &= (0.5513 * 0.5849) + (0.0372 * 0.0345) = 0.3233 \\ x_1 \cdot x_3 &= (0.5513 * 0.1164) + (0.0372 * 0.1589) = 0.0696 \\ x_1 \cdot x_4 &= (0.5513 * 0.5564) + (0.0372 * 0.1105) = 0.3164 \end{aligned}$$

Lanjutkan untuk semua elemen. Matriks kernel lengkap:

$$K = \begin{bmatrix} 0.3048 & 0.3233 & 0.0696 & 0.3164 \\ 0.3233 & 0.3423 & 0.0723 & 0.3302 \\ 0.0696 & 0.0723 & 0.0272 & 0.0928 \\ 0.3164 & 0.3302 & 0.0928 & 0.3184 \end{bmatrix}$$

Substitusi ke fungsi Dual :

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \alpha_i \alpha_j y_i y_j K_{ij} \quad (3)$$

Berdasarkan Persamaan (3) diatas, substitusi nilai $y_i (1, 1, -1, -1)$ dan K_{ij} :

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}^T \begin{bmatrix} 0.3048 & 0.3233 & -0.0696 & -0.3164 \\ 0.3233 & 0.3423 & -0.0723 & -0.3302 \\ -0.0696 & -0.0723 & 0.0272 & 0.0928 \\ -0.3164 & -0.3302 & 0.0928 & 0.3184 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

Maka hasil yang didapat adalah :

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - 0.1524 (\alpha_1^2 \alpha_2^2 \alpha_3^2 \alpha_4^2) + (\text{Interaksi antar } \alpha_i)$$

Misalkan $\alpha_1 = 0.5$, $\alpha_3 = 0.5$, dan lainnya adalah 0.

$$w = \sum_{i=1}^4 \alpha_i y_i x_i \quad (4)$$

Substitusi nilai menggunakan Persamaan (4) :

$$\begin{aligned} w &= (0.5(1)([0.5513, 0.0372]) + 0.5(-1)([0.1164, 0.1589])) \\ w &= [0.2175, -0.0609] \end{aligned}$$

Hitung b , dengan menggunakan salah satu support vector (x_1):

$$\begin{aligned} b &= y_1 - w \cdot x_1 \\ b &= 1 - ([0.2175, -0.0609] * [0.5513, 0.0372]) \\ b &= 1 - (0.1199 - 0.0023) = 0.8824 \end{aligned}$$

Fungsi keputusan yang didapat adalah $f(x) = [0.2175, -0.0609] * x + 0.8824$

Prediksi dilakukan dan mendapatkan hasil seperti pada Tabel 3 dibawah ini:

Tabel 3. Hasil pengujian SVM

Label	Data Uji		Fungsi Keputusan Metode 5-Fold Cross-Validation	y_1	Keterangan
	x_1	x_2	$f(x) = \text{sign}(w_1 \cdot x_1 + w_2 \cdot x_2 + b)$		

1	0.5782	0.0373	Sign(1.0058)	1	Kelas Ajwa
2	0.13141	0.0959	Sign(0.9058)	1	Kelas Deglet
3	0.1139	0.1652	Sign(0.8971)	1	Bukan Kelas Golden
4	0.4987	0.0781	Sign(0.9861)	1	Bukan Kelas Sukari

4.4 Penerapan Algoritma Naive Bayes

Pada penelitian ini, algoritma Gaussian Naive Bayes digunakan untuk mengklasifikasikan empat jenis kurma, yaitu Kurma Ajwa, Kurma Deglet, Kurma Golden, dan Kurma Sukari. Algoritma ini bekerja dengan mengasumsikan bahwa fitur-fitur yang digunakan memiliki distribusi normal (Gaussian).

Persamaan (5) dari Gaussian meliputi : Untuk setiap fitur x_i , distribusi probabilitas dihitung sebagai berikut:

$$P(x_i | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \quad (5)$$

Dimana :

μ : Rata-rata fitur dalam data latih untuk kelas tertentu.

σ : Standar deviasi fitur dalam data latih untuk kelas tertentu.

x_i : Nilai fitur pada data uji.

Menghitung probabilitas untuk setiap fitur:

Probabilitas untuk Mean_H:

$$P(\text{Mean}_H = 0.578192 | \mu = 0.551266, \sigma = 0.141054)$$

Kemudian menggunakan persamaan Gaussian Naive Bayes :

$$P(0.578192 | 0.551266, 0.141054) = \frac{1}{\sqrt{2\pi(0.141054)^2}} * e^{-\frac{(0.578192-0.551266)^2}{2(0.141054)^2}}$$

$$\text{Probabilitas } P(0.578192) = \frac{1}{0.3535} * 0.9819 = 2.776$$

Dan hasil selanjutnya dapat dilihat seperti pada Tabel 4.

Probabilitas	Hasil
Mean_H	2.776
Mean_S	13.04
Mean_V	0.697
Std_H	0.0775
Std_S	12.79
Std_V	2.441

Menghitung Probabilitas Total :

$$P(\text{Class 1} | \text{Data Uji 1})$$

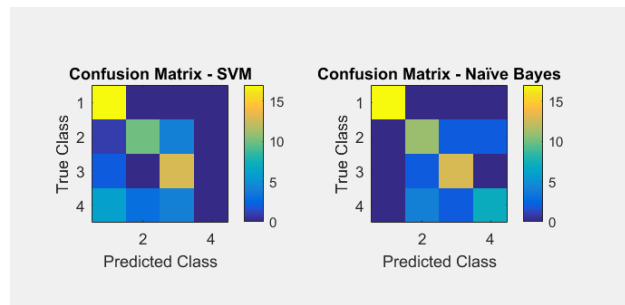
$$= P(\text{Mean}_H) * P(\text{Mean}_S) * P(\text{Mean}_V) * P(\text{Std}_H) * P(\text{Std}_S) * P(\text{Std}_V) * P(\text{Class})$$

$$P(\text{Class 1}) = 2.776 * 13.04 * 0.697 * 0.0775 * 12.79 * 2.441 = 61.03$$

Proses yang sama bisa dilakukan untuk kelas 2 yaitu (Kurma Deglet), kelas 3 yaitu (Kurma Golden) dan kurma 4 yaitu (Kurma Sukari). Kesimpulannya, proses klasifikasi menggunakan algoritma Naive Bayes untuk data uji dilakukan dengan menghitung probabilitas setiap fitur berdasarkan distribusi Gaussian untuk masing-masing kelas (Kurma Ajwa, Kurma Deglet, Kurma Golden, dan Kurma Sukari).

4.5 Perbandingan Kedua Metode

Dari Gambar 4 dapat dilihat perbandingan *confusion matrix* antara Support Vector Machine dan Naive Bayes sebagai berikut:



Gambar 4. Confusion matrix

Diketahui nilai TP, TN, FP dan FN didapat tertera pada Gambar 5 sebagai berikut :

```

--- Evaluasi Metrics ---
SVM - Kelas Kurma_Ajwa: TP=17, TN=34, FP=9, FN=0
Naïve Bayes - Kelas Kurma_Ajwa: TP=17, TN=43, FP=0, FN=0
SVM - Kelas Kurma_Deglet: TP=10, TN=42, FP=3, FN=5
Naïve Bayes - Kelas Kurma_Deglet: TP=11, TN=39, FP=6, FN=4
SVM - Kelas Kurma_Golden: TP=13, TN=37, FP=8, FN=2
Naïve Bayes - Kelas Kurma_Golden: TP=13, TN=41, FP=4, FN=2
SVM - Kelas Kurma_Sukari: TP=0, TN=47, FP=0, FN=13
Naïve Bayes - Kelas Kurma_Sukari: TP=7, TN=45, FP=2, FN=6

SVM - Precision: 68.07%, Recall: 63.33%, F1-Score: 74.24%
Naïve Bayes - Precision: 79.74%, Recall: 78.46%, F1-Score: 78.41%
    
```

Gambar 5. Evaluasi matrix

Adapun perhitungan manual dari kedua algoritma ini yaitu :

1. Algoritma SVM

$$\begin{aligned}
 accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \\
 &= \frac{(17 + 34) + (10 + 42) + (13 + 37) + (0 + 47)}{(17 + 34 + 9 + 0) + (10 + 42 + 3 + 5) + \dots + (0 + 47 + 0 + 13)} \times 100\% \\
 &= \frac{180}{60} \times 100 \\
 &= 66.67\%
 \end{aligned}$$

Kurma Ajwa :

$$precision = \frac{17}{17+9} = 0.6545 \quad Recall = \frac{17}{17+0} = 1 \quad F1 - Score = 2 * \frac{0.6545*1}{0.6545+1} = 0.7916$$

Kurma Deglet :

$$precision = \frac{10}{10+3} = 0.7692 \quad Recall = \frac{10}{10+5} = 0.667 \quad F1 - Score = 2 * \frac{0.7692*0.667}{0.7692+0.667} = 0.7143$$

Kurma Golden :

$$precision = \frac{13}{13+8} = 0.6190 \quad Recall = \frac{13}{13+2} = 0.8667 \quad F1 - Score = 2 * \frac{0.6190*0.8667}{0.6190+0.8667} = 0.7916$$

Kurma Sukari :

$$precision = \frac{0}{0+0} = 0 \quad Recall = \frac{0}{0+13} = 0 \quad F1 - Score = 2 * \frac{0*0}{0+0} = 0$$

Rata-rata :

$$Precision = \frac{0.6545 + 0.7692 + 0.6190 + 0}{4} = 0.68 \text{ atau } 68.07\%$$

$$Recall = \frac{1 + 0.667 + 0.8667 + 0}{4} = 0.63 \text{ atau } 63.33\%$$

$$F1 - Score = \frac{0.7916 + 0.7143 + 0.7234 + 0}{4} = 0.74 \text{ atau } 74.24\%$$

2. Algoritma Naïve Bayes

$$\begin{aligned}
 accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \times 100 \\
 &= \frac{(17 + 43) + (11 + 39) + (13 + 41) + (7 + 45)}{(17 + 34 + 9 + 0) + (10 + 42 + 3 + 5) + \dots + (0 + 47 + 0 + 13)} \times 100\%
 \end{aligned}$$

$$= \frac{240}{60} \times 100$$

$$= 80\%$$

Kurma Ajwa :

$$precision = \frac{17}{17+0} = 1 \quad Recall = \frac{17}{17+0} = 1 \quad F1 - Score = 2 * \frac{1*1}{1+1} = 1$$

Kurma Deglet :

$$precision = \frac{11}{11+6} = 0.647 \quad Recall = \frac{11}{11+4} = 0.733 \quad F1 - Score = 2 * \frac{0.647*0.733}{0.647+0.733}$$

$$= 0.687$$

Kurma Golden :

$$precision = \frac{13}{13+4} = 0.765 \quad Recall = \frac{13}{13+2} = 0.8667 \quad F1 - Score = 2 * \frac{0.765*0.8667}{0.765+0.8667}$$

$$= 0.813$$

Kurma Sukari :

$$precision = \frac{7}{7+2} = 0.778 \quad Recall = \frac{7}{7+6} = 0.538 \quad F1 - Score = 2 * \frac{0.778*0.538}{0.778+0.538}$$

$$= 0.635$$

Rata-rata :

$$Precision = \frac{1 + 0.647 + 0.765 + 0.778}{4} = 0.798 \text{ atau } 79.8\%$$

$$Recall = \frac{1 + 0.733 + 0.8667 + 0.538}{4} = 0.785 \text{ atau } 78.5\%$$

$$F1 - Score = \frac{1 + 0.687 + 0.813 + 0.635}{4} = 0.784 \text{ atau } 78.4\%$$

Analisis *confusion matrix* dilakukan untuk memahami distribusi prediksi model terhadap masing-masing kelas. Visualisasi *confusion matrix* disajikan pada gambar diartas untuk memberikan informasi lebih lanjut tentang klasifikasi yang berhasil dan kesalahan klasifikasi model. Berikut output yang ditampilkan terlihat pada Gambar 6.

```
Jumlah data latih: 140
Jumlah data uji: 60
Akurasi SVM: 66.67%
Akurasi Naïve Bayes: 80.00%
Naïve Bayes memiliki performa lebih baik.
```

Gambar 6. Hasil akurasi uji

Berdasarkan pengujian, akurasi metode SVM adalah 66,67%, sedangkan akurasi metode Naïve Bayes mencapai 80,00%, menunjukkan bahwa Naïve Bayes memiliki performa yang lebih **baik** dibandingkan SVM dalam klasifikasi dataset ini. Hal ini dapat disebabkan oleh karakteristik dataset yang lebih cocok untuk pendekatan probabilistik seperti Naïve Bayes. Dan Gambar 7 dibawah merupakan output bahwasanya data yang digunakan pada penelitian tidak ada overfitting.

```
--- Evaluasi Overfitting ---
Akurasi SVM (Latih): 60.71%
Akurasi Naïve Bayes (Latih): 76.43%
Model tidak mengalami overfitting.
fx >>
```

Gambar 7. Hasil akurasi latih

Gambar tersebut menampilkan hasil evaluasi model SVM dan Naïve Bayes terhadap data latih. Akurasi model SVM pada data latih tercatat sebesar 60,71%, sedangkan akurasi model Naïve Bayes mencapai 76,43%. Selain itu, evaluasi menyimpulkan bahwa model tidak mengalami overfitting, yang berarti performa model pada data latih dan data uji cenderung seimbang, tanpa indikasi model terlalu menyesuaikan data latih sehingga mengorbankan generalisasi pada data baru.

5 Kesimpulan

Proses klasifikasi menggunakan algoritma SVM dengan *kernel linier* menghasilkan akurasi sebesar 66,67%, sedangkan algoritma *Gaussian Naïve Bayes* mencapai akurasi sebesar 80,00%. Hasil

ini menunjukkan bahwa Naïve Bayes memiliki performa yang lebih baik dalam mengklasifikasikan jenis kurma dibandingkan SVM. Evaluasi *overfitting* menunjukkan bahwa model tidak mengalami *overfitting*, dengan performa yang konsisten antara data latih dan data uji. Perbandingan performa antara algoritma SVM dan *Naive Bayes* dalam klasifikasi buah kurma menunjukkan bahwa *Naive Bayes* memiliki kinerja yang lebih baik secara keseluruhan. Selain itu, *metrik precision, recall, dan F1-score Naive Bayes* juga lebih unggul, masing-masing sebesar 79.74%, 78.46%, dan 79.09%, dibandingkan dengan SVM yang hanya mencapai 52.49%, 65.00%, dan 58.08%.

6 Referensi

- [1] M. Giamouzi, "City , University of London Institutional Repository," vol. 34, no. 2019, pp. 51–79, 2008, doi: 10.1007/978-3-030-69717-4.
- [2] M. Fandi, "Aplikasi Identifikasi Jenis Buah Kurma dengan Metode GLCM Berbasis Android," *J. Pengemb. Rekayasa dan Teknol.*, vol. 16, no. 1, p. 34, 2020, doi: 10.26623/jprt.v16i1.2109.
- [3] R. Risna, F. Amanda, and S. Uyun, "Comparison of K-Nearest Neighbor (KNN) and Linear Discriminant Analysis (LDA) Algorithms for Mature Ajwa Date Fruit Classification," *Int. Conf. Inf. Sci. Technol. Innov.*, vol. 2, no. 1, pp. 11–16, 2023, doi: 10.35842/icostec.v2i1.29.
- [4] M. H. S. Ajat, "Klasifikasi SMS Spam dengan Komparasi Metode SVM dan Naïve Bayes," *Method. J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 1, pp. 31–34, 2023, doi: 10.46880/mtk.v9i1.1694.
- [5] R. Fazal, "Membandingkan Support Vector Machines dan Naive Bayes pada Analisis Sentimen Data Twitter," *J. Portal Data*, vol. 2, no. 10, pp. 1–9, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/248%0Ahttp://portaldata.org/index.php/portaldata/article/download/248/218>
- [6] Ellif, S. H. Sitorus, and R. Hidayati, "Naïve Bayes 1.," *Coding J. Komput. dan Apl.*, vol. 09, no. 01, pp. 66–75, 2021.
- [7] A. A. Muhammad, A. Arkadia, S. NaufalRifqi, Trianto, and D. S. Prasvita, "Klasifikasi Tingkat Kematangan Buah Pisang berdasarkan Fitur Warna dengan Metode SVM," *J. Ilmu Komput. dan Desain Komun. Vis.*, vol. 4, no. 1, pp. 9–16, 2019.
- [8] K. Ayuningsih, Y. A. Sari, and P. P. Adikara, "Klasifikasi Citra Makanan menggunakan HSV Color Moment dan LBP dengan Naïve Bayes Classifier," *J. Pengemb. Teknol. Informasii dan Ilmu Komputier Univ. Brawijaya*, vol. 3, no. 4, pp. 3166–3173, 2019.
- [9] Milda, D. M. Sari, and M. Rafli, "Klasifikasi Kematangan Buah Pisang Kepok berdasarkan Citra HSV dengan K-Nearest Neighbors," *Pros. Semin. Nas. Rekayasa Keteknikan Inform. (SENARAI 2023)*, vol. 1, no. 1, pp. 1–11, 2023.
- [10] M. Afriansyah, J. Saputra, Y. Sa'adati, and V. Yoga P. Ardhana, "Optimasi Algoritma Naive Bayes untuk Klasifikasi Buah Apel berdasarkan Fitur Warna RGB," *Bull. Comput. Sci. Res.*, vol. 3, no. 3, pp. 242–249, 2023, doi: 10.47065/bulletincsr.v3i3.251.
- [11] M. R. Saputra and H. Irsyad, "Klasifikasi Tingkat Kemanisan Alpukat berdasarkan Fitur Hue Saturation Value (HSV) dengan menggunakan Support Vector Machine (SVM)," *J. Algoritm.*, vol. 2, no. 2, pp. 113–119, 2022, doi: 10.35957/algoritme.v2i2.2361.
- [12] J. Sisfotenika, "Implementasi Support Vector Machine (SVM) pada Klasifikasi Jenis Tanah memanfaatkan Fitur RGB Implementation of SVM in Soil Type Classification using RGB Features," vol. 14, no. 2, pp. 175–184, 2024.
- [13] M. Afriansyah, J. Saputra, V. Y. P. Ardhana, and Yuan Sa'adati, "Algoritma Naive Bayes yang Efisien untuk Klasifikasi Buah Pisang Raja berdasarkan Fitur Warna," *J. Inf. Syst. Manag. Digit. Bus.*, vol. 1, no. 2, pp. 236–248, 2024, doi: 10.59407/jismdb.v1i2.438.
- [14] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes dalam memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- [15] T. Widyanto, I. Ristiana, and A. Wibowo, "Komparasi Naive Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 3, pp. 147–161, 2023, doi: 10.31598/sintechjournal.v6i3.1433.